W358–W363 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* *Published online 16 May 2008*
doi:10.1093/nar/gkn276

GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis

Qi Zheng[1,2] and Xiu-Jie Wang[1,*]

[1]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences and [2]Graduate University of the Chinese Academy of Sciences, Beijing 100101, China

Received January 30, 2008; Revised April 7, 2008; Accepted April 26, 2008

ABSTRACT

Gene Ontology (GO) analysis has become a commonly used approach for functional studies of large-scale genomic or transcriptomic data. Although there have been a lot of software with GO-related analysis functions, new tools are still needed to meet the requirements for data generated by newly developed technologies or for advanced analysis purpose. Here, we present a Gene Ontology Enrichment Analysis Software Toolkit (GOEAST), an easy-to-use web-based toolkit that identifies statistically overrepresented GO terms within given gene sets. Compared with available GO analysis tools, GOEAST has the following improved features: (i) GOEAST displays enriched GO terms in graphical format according to their relationships in the hierarchical tree of each GO category (biological process, molecular function and cellular component), therefore, provides better understanding of the correlations among enriched GO terms; (ii) GOEAST supports analysis for data from various sources (probe or probe set IDs of Affymetrix, Illumina, Agilent or customized microarrays, as well as different gene identifiers) and multiple species (about 60 prokaryote and eukaryote species); (iii) One unique feature of GOEAST is to allow cross comparison of the GO enrichment status of multiple experiments to identify functional correlations among them. GOEAST also provides rigorous statistical tests to enhance the reliability of analysis results. GOEAST is freely accessible at http://omicslab.genetics.ac.cn/GOEAST/

INTRODUCTION

High-throughput experimental techniques, such as microarray and mass spectrometry, have become indispensable tools for many biological studies in the postgenomic era. However, such high-throughput experiments often generate hundreds of candidate genes or proteins, sometimes with noisy results. How to study the functional relationships among selected genes/proteins and how to identify the key-regulatory components have therefore become important issues for interpreting high-throughput experimental data.

As a uniform knowledge base, Gene Ontology (GO) provides controlled vocabulary to describe gene products in three categories, namely biological process, molecular function and cellular component (1). It has become a commonly used resource for gene function studies. Identification of overpresented GO terms among a given list of genes could help biologists to better understand the functional relevance of these genes.

There have been a number of tools that are capable of carrying out GO term enrichment analysis, some of which also have pathway analysis or other functions (1–28). However, a considerable number of these tools require local installation, sometimes on particular operating systems, which makes them inconvenient to use. Some tools use statistical methods that have been proven to be inappropriate for GO term enrichment analysis in small datasets (29), or use out-of-date database to calculate the background GO term distribution, therefore, may render the accuracy of the analysis (2,3,11,12,18). In addition, the appearance of new high-throughput experimental technologies and the increasingly complicated experimental design also raised the needs for new GO analysis tools with novel functions. For example, microarrays manufactured by Illumina and Agilent incorporations have attracted increasing number of users, but few available tools support analysis for direct use of probe IDs of these gene chips yet. There are needs for comparison of the GO enrichment status from different experiments as well. In addition, most available tools only provide a list of enriched GO terms as analysis results, without showing the correlations among these GO terms (2–6,8–11,15,16,18,22,24–27). However, as GO categories have hierarchically organized tree structures, knowing the relationships of different GO terms would help biologists to better understand their experimental results from the functional point of view.

*To whom correspondence should be addressed. Tel: +86 10 64840941; Fax: +86 10 64873428; Email: xjwang@genetics.ac.cn

To solve the aforementioned problems/shortcomings of available GO analysis tools, we developed GOEAST, a Gene Ontology Enrichment Analysis Software Toolkit. GOEAST is a web-based user friendly tool, which applies appropriate statistical methods to identify significantly enriched GO terms among a given list of genes. GOEAST supports analysis for data of about 60 species and from various resources, including probe or probe set IDs of Affymetrix, Illumina, Agilent or customized microarrays and gene identifiers from various databases. It also provides graphical outputs of enriched GO terms to demonstrate their relationships in the three ontology categories. A unique feature of GOEAST is to support cross-comparisons of the GO enrichment status of multiple experiments, therefore to identify the correlations and differences among them.

MATERIALS AND METHODS

Input format

GOEAST requires text-format input of a list of gene identifiers. The gene identifiers supported by GOEAST include probe set IDs of Affymetrix gene chips, target names, search keys or probe IDs of Illumina microarrays, probe IDs of Agilent microarrays and gene/protein IDs of various databases, such as NCBI, RefSeq, Ensembl, UniProtKB, Sanger GeneDB, MGI, RGD, FlyBase, WormBase, TAIR, Gramene, etc.

Data resources

Data used by GOEAST were obtained from the following resources. The GO ontology files were downloaded from the website of the Gene Ontology Project, both the definition and hierarchical relationships of all GO terms were extracted from the ontology files by Perl scripts. The annotation files of Affymetrix gene chips, Illumina microarrays and Agilent microarrays were downloaded from the companies' websites, respectively. The associated GO terms of probes on all commercial microarrays were extracted from the corresponding array annotation files. The GO annotation files for non-microarray data were downloaded from the Gene Ontology ftp server and parsed by Perl scripts to extract associated GO terms (Supplementary Material 1).

Identification of enriched GO terms

The default statistical method used by GOEAST to identify significantly enriched GO terms among a given list of genes is hypergeometric test. If an input file contains valid IDs of k genes from a microarray with a total of t genes, for a given GO term, there are q genes within k and m genes within t associated with it, then the possibility that whether genes associated with this GO term is enriched among the input gene list could be calculated by hypergeometric test,

$$P(X = x > q) = \sum_{x=q}^{m} \frac{\binom{m}{x}\binom{t-m}{k-x}}{\binom{t}{k}}.$$

GOEAST also supports Fisher's exact test and χ^2-test. In addition, GOEAST also provides the log-odds ratio (LR) of genes associated with each GO term in the input file and the entire array, which is calculated as:

$$\mathrm{LR} = \log_2(q/k)/(m/t) = \log_2(q/k) - \log_2(m/t).$$

For any GO term, genes directly belong to it as well as those belong to any of its offspring GO terms are all considered as its associated genes. If gene identifiers instead of probe/probe set IDs are given as input, the GO information of all annotated genes of that species will be used to calculate m and t. All statistical analysis is carried out using R packages. Since the statistical analysis is not appropriate to questions with small sample size, GO terms with m smaller than 5 were discarded in the analysis. To circumvent the multi-test problem which might induce too much false positive results, GOEAST by default adjusts the raw P-values into false discovery rate (FDR) using the Benjamini–Yekutieli method (30). Users can also choose other methods for multi-test problem correction or not to use multi-test problem correction at all.

Outputs

GOEAST provides three types of output files, an HTML table of enriched GO terms which provides detailed information as well as AmiGO link for enriched GO terms and their associated genes, a plain-text file of enriched GO terms for local processing and Multi-GOEAST analysis, and three graphical outputs showing the hierarchical relationships of enriched GO terms in each GO category. Graphs are generated by the Graphviz software. Multiple formats, including PDF, PostScript, SVG, GIF, JPEG and PNG formats, are supported for the graphical output files. When generating graphs, GOEAST uses different color saturation degrees to present the enrichment significance of each GO term. The saturation degree S for the background color of the box representing a GO term is defined as $S \propto -\log P$, where S ranges from 0 to 1. In the graphical output of Multi-GOEAST results, combination of colors for individual experiment is used as the color set of enriched GO terms shared by multiple experiments. The saturation degrees of boxes representing common GO terms are determined by the lowest P-value of each GO term in all experiments. Grayscale is used for boxes representing GO terms shared by three experiments, for which the brightness level instead of saturation level varies according to P-values.

Implementations and updates

GOEAST is constructed and configured upon a typical LAMP (Linux + Apache + MySQL + PHP) platform. The supporting Apache2 server runs on a GUN/Linux operating system. Precomputed data are stored in MySQL 5.0 DBMS. All static content of the website is written in XHTML language, processing and database querying functionalities are achieved by PHP5 scripts. All GOEAST web pages were constructed following the W3C recommended XHTML1.0 Transitional DTD standard, which makes GOEAST compatible with most available web browsers.

GOEAST uses most up-to-date source data for analysis. Source data files are downloaded automatically or semi-automatically to the GOEAST server according to the update rate of each file. For example, the annotation files of Affymetrix microarrays will be updated every 4 months and the GO ontology files will be updated every week (see Supplementary Material 1 for details). GOEAST will remain available and continue to update for at least 5 years after release.

Functionalities and features

The main function of GOEAST is to identify statistically significantly enriched GO terms among a given list of genes. As a web-based GO enrichment analysis tool, GOEAST has the following improved or unique features compared with available tools.

Broadness of support. Most available GO analysis tools only support gene identifiers in the format of Affymetrix probe set IDs, NCBI IDs or identifiers from other major sequence databases. However, with increasing number of data generated by Illumina and Agilent microarrays, and the broader application of other high-throughput technologies, more and more users need to carry out additional works when the format of their gene identifiers are incompatible with the requirement of available tools. To overcome such problem, GOEAST provides supports for probe identifiers of all available Affymetrix, Illumina and Agilent microarrays, as well as gene identifiers from various databases (Supplementary Material 1). A total of ~60 prokaryote and eukaryote species are supported by GOEAST, a number much higher than that supported by most other similar tools. Please see Supplementary Material 2 for a detailed comparison of GOEAST with similar tools.

Graphical outputs. The GO terms within each ontology category are not independent but located in the same vocabulary tree with hierarchical relationships to one another. Knowing the positional relationships of enriched GO terms will help users to better understand their results. Figure 1 shows the GOEAST analysis results of a sample list of genes. With the graphical view, one can easily tell that many transcription-related GO terms and their children terms are enriched. On the other hand, the overpresentation of GO term ‘growth’ might be caused by various factors because none of its offspring terms is found to be enriched. In the graphical output, GOEAST displays the *P*-value of each enriched term and uses different color saturation degrees to present the enrichment significance of different GO terms, which make the results easily to be understood. Among the few available GO analysis tools with graphical output (12,19,20,23,28), only EasyGO has similar functions, but it is limited to several plants and farm animals (28).

Multiple experiment comparison function. One unique feature of GOEAST is to allow comparison of GO term enrichment status in multiple experiments. Users can upload the GO term enrichment analysis results provided by GOEAST microarray or batch-gene tools of two or three experiments to the Multi-GOEAST website to identify the commonly and specifically enriched GO terms of each experiment, thereby to discover information missed by analyzing each experiment alone. For example, in a work studying the functions of two histone demethylase genes (31), *Jmjd1a* and *Jmjd2c*, the authors used Illumina microarrays to examine the gene expression profile changes in germ cells of *Jmjd1a* and *Jmjd2c* knockdown mice, respectively, and identified a couple of downstream genes regulated by *Jmjd1a* or *Jmjd2c*. But the paper did not provide further comparison of these *Jmjd1a* and *Jmjd2c* downstream genes. We first used GOEAST to analyze the GO enrichment status of the reported differentially expressed genes in *Jmjd1a* and *Jmjd2c* knockdown mice, respectively, then compared these two GO enrichment analysis results using the Multi-GOEAST function, and found that although both *Jmjd1a* and *Jmjd2c* can regulate gene transcription, *Jmjd2c* is more involved in negative regulatory processes (Figure 2). This result is consistent with the authors’ finding that *Jmjd2c* regulates the binding of some transcription corepressors to Nanog, a key transcription factor for embryonic stem cell development. In addition, Multi-GOEAST analysis also found other interesting phenomena that were not reported by the authors, such as metabolic process-related GO terms, especially glycolysis process-related GO terms, are more enriched among differentially expressed genes found in the *Jmjd2c* knockdown mice, whereas development-related GO terms are more enriched among differentially expressed genes found in the *Jmjd1a* knockdown mice.

DISCUSSION

GOEAST is an online GO term enrichment analysis tool. It is developed with improved functionalities to meet new needs appeared with the broad application of multiple high-throughput experimental technologies, such as microarray, SAGE, mass spectrometry, etc.

The web-based nature makes GOEAST very easy to use; analysis can be finished with a few clicks. The source data files used by GOEAST are automatically or semi-automatically updated, which ensures that the users will always receive the most up-to-date analysis results. The broadness of supported microarray types and species allows more users to take advantage of GOEAST. As have shown by aforementioned examples, the graphical output and multiple experiment comparison function of GOEAST are capable of identifying more useful information from input data.

Some available GO enrichment analysis tools use a universal gene set or the input gene list itself as the statistical background for calculating GO term enrichment *P*-values (2,3,11,12,18), which would produce biased enrichment analysis results. To overcome this problem, GOEAST uses all probes in each microarray platform or all genes of a given species to calculate the background GO distribution, thus ensures the accuracy of analysis results. The commonly used statistical methods for GO enrichment analysis include binomial test, χ^2-test, Fisher’s exact test and hypergeometric test (Supplementary

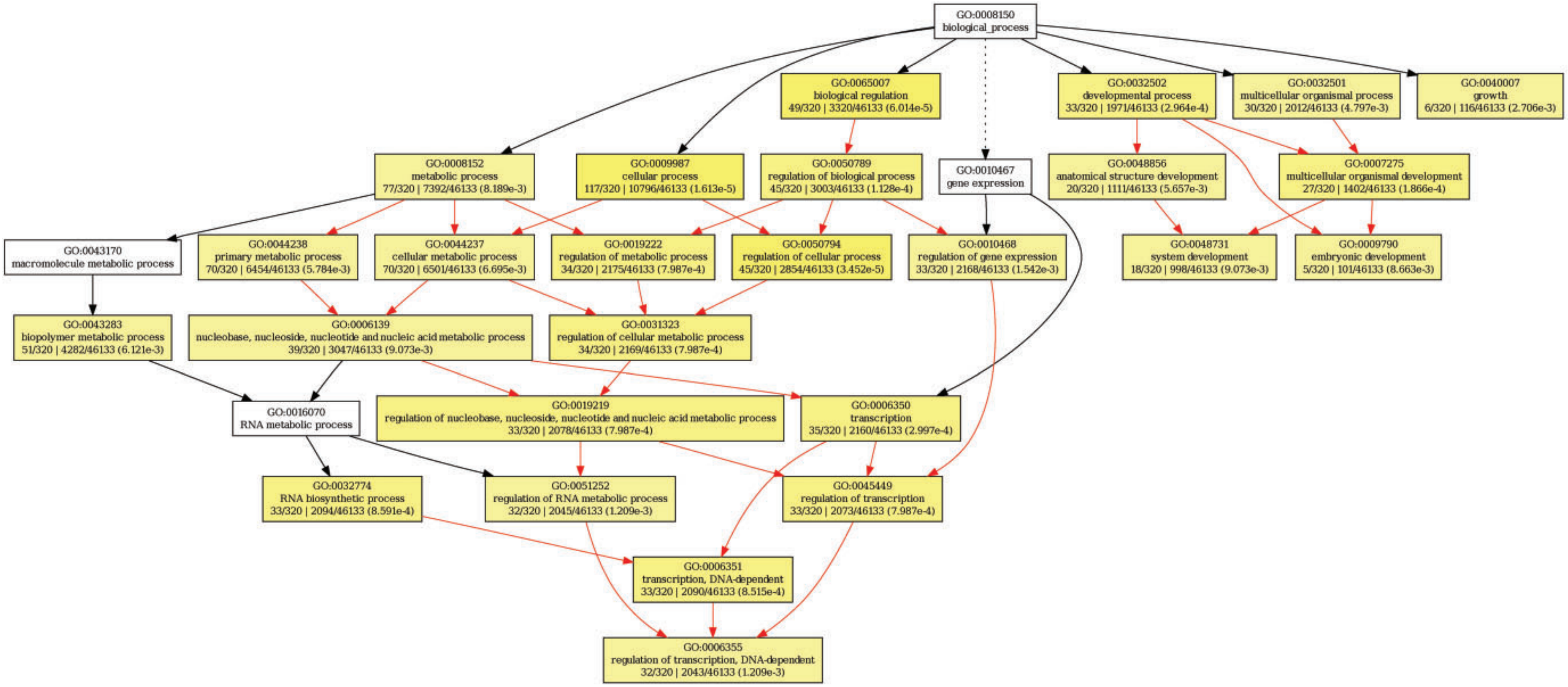
GO:0008150
biological_process
GO:0065007
biological regulation
49/320 | 3320/46133 (6.014e-5)
GO:0032502
developmental process
33/320 | 1971/46133 (2.964e-4)
GO:0032501
multicellular organismal process
30/320 | 2012/46133 (4.797e-3)
GO:0040007
growth
6/320 | 116/46133 (2.706e-3)
GO:0008152
metabolic process
77/320 | 7392/46133 (8.189e-3)
GO:0009987
cellular process
117/320 | 10796/46133 (1.613e-5)
GO:0050789
regulation of biological process
45/320 | 3003/46133 (1.128e-4)
GO:0010467
gene expression
GO:0048856
anatomical structure development
20/320 | 1111/46133 (5.657e-3)
GO:0007275
multicellular organismal development
27/320 | 1402/46133 (1.866e-4)
GO:0043170
macromolecule metabolic process
GO:0044238
primary metabolic process
70/320 | 6454/46133 (5.784e-3)
GO:0044237
cellular metabolic process
70/320 | 6501/46133 (6.695e-3)
GO:0019222
regulation of metabolic process
34/320 | 2175/46133 (7.987e-4)
GO:0050794
regulation of cellular process
45/320 | 2854/46133 (3.452e-5)
GO:0010468
regulation of gene expression
33/320 | 2168/46133 (1.542e-3)
GO:0048731
system development
18/320 | 998/46133 (9.073e-3)
GO:0009790
embryonic development
5/320 | 101/46133 (8.663e-3)
GO:0043283
biopolymer metabolic process
51/320 | 4282/46133 (6.121e-3)
GO:0006139
nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
39/320 | 3047/46133 (9.073e-3)
GO:0031323
regulation of cellular metabolic process
34/320 | 2169/46133 (7.987e-4)
GO:0016070
RNA metabolic process
GO:0019219
regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process
33/320 | 2078/46133 (7.987e-4)
GO:0006350
transcription
35/320 | 2160/46133 (2.997e-4)
GO:0032774
RNA biosynthetic process
33/320 | 2094/46133 (8.591e-4)
GO:0051252
regulation of RNA metabolic process
32/320 | 2045/46133 (1.209e-3)
GO:0045449
regulation of transcription
33/320 | 2073/46133 (7.987e-4)
GO:0006351
transcription, DNA-dependent
33/320 | 2090/46133 (8.515e-4)
GO:0006355
regulation of transcription, DNA-dependent
32/320 | 2043/46133 (1.209e-3)

Figure 1. The GOEAST graphical output of enriched GO terms in the biological process category for a sample experiment. Boxes represent GO terms, labeled by its GO ID, term definition, and detailed information, organized as '$q/m|t/k$ (P-value)' (see Materials and Methods section for the meanings of q, m, t and k). Significantly enriched GO terms are marked yellow. The degree of color saturation of each node is positively correlated with the enrichment significance of the corresponding GO term. Nonsignificant GO terms within the hierarchical tree are shown as white boxes. Branches of the GO hierarchical tree without significantly enriched GO terms are not shown. Arrows represent connections between different GO terms. Red arrows represent relationships between two enriched GO terms, black solid arrows represent relationships between enriched and unenriched terms and black dashed arrows represent relationships between two unenriched GO terms.

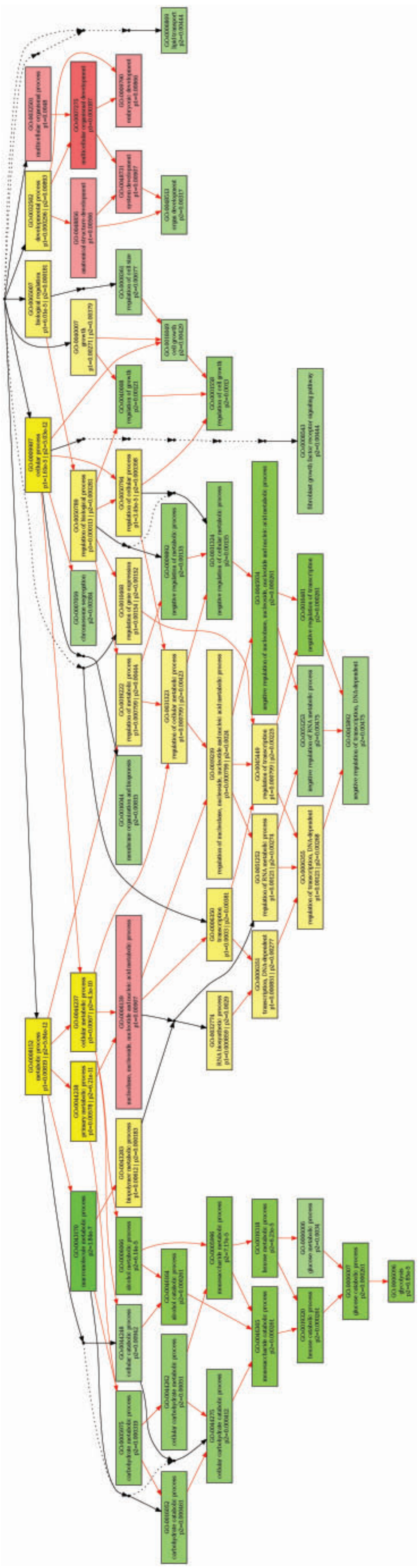

Figure 2. Graphical output of Multi-GOEAST analysis results for differentially expressed genes in *Jmjd1a* and *Jmjd2c* knockdown mice. Red and green boxes represent enriched GO terms only found in *Jmjd1a* knockdown mice (first experiment) or *Jmjd2c* knockdown mice (second experiment), respectively. Yellow boxes represent commonly enriched GO terms in both experiments. The saturation degrees of all colors represent the significance of enrichment for corresponding GO terms.

Table 3). Previous research has shown that Fisher's exact test and hypergeometric test are more accurate than binomial test and χ^2-test for GO enrichment analysis, for the latter two tests require large sample sizes that most GO terms do not have (29). Since the Fisher's exact test and the hypergeometric test are statistical equivalent (29), GOEAST uses hypergeometric test by default because it is much faster to be calculated than Fisher's exact test by R packages. Fisher's exact test and χ^2-test are also supported by GOEAST.

When doing the same statistical test many times, the multiple-test problem often becomes significant and will produce more false positive results. This problem is usually solved by controlling the overall FDR of the entire result instead of controlling type I errors (the raw *P*-values) in each individual test or controlling the family-wise error rate (FWER), which is considered to be too strict for biological data (32). There are two commonly used methods to control FDRs, the Benjamini–Hochberg method and the Benjamini–Yekutieli method (33). The former one is suitable for independent multiple-tests whereas the later is suitable for positively related multiple-tests. Since the enriched GO terms among a given list of genes are often positively related, we chose the Benjamini–Yekutieli method to calculate FDRs in GOEAST by default. Several other adjustment methods are also supported.

Due to the hierarchical-dependent relationships of GO terms, the enrichment of some GO terms might also cause overpresentation of their neighboring terms, when the correlations of neighboring GO terms and their enrichment status were considered in the calculation. To overcome this problem, Alexa *et al.* (34) introduced an improved weight scoring algorithm to calculate the significance of GO terms that is thought to be able to reduce FDRs caused by overpresentation of neighboring GO terms. This algorithm is also supported by GOEAST, but since analysis using this algorithm is time consuming, the algorithm is not chosen as default method, users can activate it via the advanced parameter setting choice.

All functionalities of GOEAST have been tested thoroughly by different input data on various operating systems. We believe the development of GOEAST would help more biologists to discover hidden information of their high-throughput experimental results.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Mr Luc Jouneau and Dr Zhongming Han for their suggestive discussions. We thank Dr Meng Wang and Ms Zhongqin Yang for data collection. This project was supported by grants from the National Natural Science Foundation of China (30725014 and 90612019), the Ministry of Science and Technology of China (2007CB946901) and the Chinese Academy of Sciences (KSCX2-YW-R-134 and KSCX2-YW-N-024).

Funding to pay the Open Access publication charges for this article was provided by the National Natural Science Foundation of China.

Conflict of interest statement. None declared.

REFERENCES

1. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
2. Khatri,P., Draghici,S., Ostermeier,G.C. and Krawetz,S.A. (2002) Profiling gene expression using onto-express. *Genomics*, **79**, 266–270.
3. Robinson,M.D., Grigull,J., Mohammad,N. and Hughes,T.R. (2002) FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinform.*, **3**, 35.
4. Berriz,G.F., King,O.D., Bryant,B., Sander,C. and Roth,F.P. (2003) Characterizing gene sets with FuncAssociate. *Bioinformatics*, **19**, 2502–2504.
5. Castillo-Davis,C.I. and Hartl,D.L. (2003) GeneMerge—post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics*, **19**, 891–892.
6. Dennis,G. Jr, Sherman,B.T., Hosack,D.A., Yang,J., Gao,W., Lane,H.C. and Lempicki,R.A. (2003) DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.*, **4**, P3.
7. Draghici,S., Khatri,P., Bhavsar,P., Shah,A., Krawetz,S.A. and Tainsky,M.A. (2003) Onto-tools, the toolkit of the modern biologist: onto-express, onto-compare, onto-design and onto-translate. *Nucleic Acids Res.*, **31**, 3775–3781.
8. Hosack,D.A., Dennis,G. Jr, Sherman,B.T., Lane,H.C. and Lempicki,R.A. (2003) Identifying biological themes within lists of genes with EASE. *Genome Biol.*, **4**, R70.
9. Zeeberg,B.R., Feng,W., Wang,G., Wang,M.D., Fojo,A.T., Sunshine,M., Narasimhan,S., Kane,D.W., Reinhold,W.C., Lababidi,S. *et al.* (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.*, **4**, R28.
10. Al-Shahrour,F., Diaz-Uriarte,R. and Dopazo,J. (2004) FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, **20**, 578–580.
11. Beissbarth,T. and Speed,T.P. (2004) GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics*, **20**, 1464–1465.
12. Boyle,E.I., Weng,S., Gollub,J., Jin,H., Botstein,D., Cherry,J.M. and Sherlock,G. (2004) GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, **20**, 3710–3715.
13. Boyle,J. (2004) SeqExpress: desktop analysis and visualization tool for gene expression experiments. *Bioinformatics*, **20**, 1649–1650.
14. Cheng,J., Sun,S., Tracy,A., Hubbell,E., Morris,J., Valmeekam,V., Kimbrough,A., Cline,M.S., Liu,G., Shigeta,R. *et al.* (2004) NetAffx Gene Ontology Mining Tool: a visual approach for microarray data analysis. *Bioinformatics*, **20**, 1462–1463.
15. Martin,D., Brun,C., Remy,E., Mouren,P., Thieffry,D. and Jacq,B. (2004) GOToolBox: functional analysis of gene datasets based on Gene Ontology. *Genome Biol.*, **5**, R101.
16. Masseroli,M., Martucci,D. and Pinciroli,F. (2004) GFINDer: Genome Function INtegrated discoverer through dynamic annotation, statistical analysis, and mining. *Nucleic Acids Res.*, **32**, W293–W300.
17. Pasquier,C., Girardot,F., Jevardat de Fombelle,K. and Christen,R. (2004) THEA: ontology-driven analysis of microarray data. *Bioinformatics*, **20**, 2636–2643.
18. Shah,N.H. and Fedoroff,N.V. (2004) CLENCH: a program for calculating Cluster ENriCHment using the Gene Ontology. *Bioinformatics*, **20**, 1196–1197.
19. Zhang,B., Schmoyer,D., Kirov,S. and Snoddy,J. (2004) GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. *BMC Bioinform.*, **5**, 16.
20. Zhong,S., Storch,K.F., Lipan,O., Kao,M.C., Weitz,C.J. and Wong,W.H. (2004) GoSurfer: a graphical interactive tool for comparative analysis of large gene sets in Gene Ontology space. *Appl. Bioinform.*, **3**, 261–264.
21. Ben-Shaul,Y., Bergman,H. and Soreq,H. (2005) Identifying subtle interrelated changes in functional gene categories using continuous measures of gene expression. *Bioinformatics*, **21**, 1129–1137.
22. Lee,H.K., Braynen,W., Keshav,K. and Pavlidis,P. (2005) ErmineJ: tool for functional analysis of gene expression data sets. *BMC Bioinform.*, **6**, 269.
23. Maere,S., Heymans,K. and Kuiper,M. (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.
24. Newman,J.C. and Weiner,A.M. (2005) L2L: a simple tool for discovering the hidden significance in microarray expression data. *Genome Biol.*, **6**, R81.
25. Young,A., Whitehouse,N., Cho,J. and Shaw,C. (2005) OntologyTraverser: an R package for GO analysis. *Bioinformatics*, **21**, 275–276.
26. Al-Shahrour,F., Minguez,P., Tarraga,J., Medina,I., Alloza,E., Montaner,D. and Dopazo,J. (2007) FatiGO +: a functional profiling tool for genomic data. Integration of functional annotation, regulatory motifs and interaction data with microarray experiments. *Nucleic Acids Res.*, **35**, W91–W96.
27. Carmona-Saez,P., Chagoyen,M., Tirado,F., Carazo,J.M. and Pascual-Montano,A. (2007) GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol.*, **8**, R3.
28. Zhou,X. and Su,Z. (2007) EasyGO: Gene Ontology-based annotation and functional enrichment analysis tool for agronomical species. *BMC Genomics*, **8**, 246.
29. Rivals,I., Personnaz,L., Taing,L. and Potier,M.C. (2007) Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, **23**, 401–407.
30. Benjamini,Y. and Yekutieli,D. (2001) The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.*, **29**, 1165–1188.
31. Loh,Y.H., Zhang,W., Chen,X., George,J. and Ng,H.H. (2007) Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.*, **21**, 2545–2557.
32. Yang,Q., Cui,J., Chazaro,I., Cupples,L.A. and Demissie,S. (2005) Power and type I error rate of false discovery rate approaches in genome-wide association studies. *BMC Genet.*, **6(Suppl 1)**, S134.
33. Shaffer,J.P. (2007) Controlling the false discovery rate with constraints: the Newman-Keuls test revisited. *Biom. J.*, **49**, 136–143.
34. Alexa,A., Rahnenfuhrer,J. and Lengauer,T. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600–1607.